

Best-effort networks: modeling and performance analysis via large networks asymptotics

Guy Fayolle, Arnaud de La Fortelle, Jean-Marc Lasgouttes, Laurent Massoulié, James Roberts

Abstract—In this paper we introduce a class of Markov models, termed **best-effort networks**, designed to capture performance indices such as mean transfer times in data networks with best-effort service. We introduce the so-called **min bandwidth sharing policy** as a conservative approximation to the classical **max-min policy**. We establish necessary and sufficient ergodicity conditions for best-effort networks under the min policy. We then resort to the mean field technique of statistical physics to analyze network performance deriving fixed point equations for the stationary distribution of large symmetrical best-effort networks. A specific instance of such networks is the star-shaped network which constitutes a plausible model of a network with an overprovisioned backbone. Numerical and analytical study of the equations allows us to state a number of qualitative conclusions on the impact of traffic parameters (link loads) and topology parameters (route lengths) on mean document transfer time.

Keywords—best-effort service, max-min fairness, min policy, mean field, star-shaped network.

I. INTRODUCTION

Consider a network handling data flows from several users, and assume no quality of service commitments (such as minimum bandwidth allocations) have been made by the network to the users. Such a situation has been prevalent in the Internet until now, and is likely to remain so for another few years.

The preferred service model in that situation, known as best effort service, consists in allocating a fair proportion of bandwidth to contending users; see, e.g., Bertsekas and Gallager [1]. There are actually several possible notions of fairness available for this bandwidth allocation problem (see, e.g., Mo and Walrand [2] for a parametric family of fairness criteria covering all other notions proposed so far), although the classical notion proposed in [1] is the so-called max-min fairness.

Recent work has led to a relatively good understanding of how bandwidth is shared between network users when a given congestion control algorithm is used; see, e.g., Massoulié and Roberts [3] and references therein. The question of what type of fairness is achieved in the current Internet, where Jacobson's congestion avoidance algorithm—as implemented in TCP—is responsible for congestion control, has been studied in depth by Hurley et al. [4]. These studies all assume the number of flows remains fixed.

In comparison, there is little work accounting for the random nature of traffic and its impact on user perceived quality of service. Consider for instance the transfer of digital documents (Web pages, files, emails,...) using a transport protocol like TCP. This constitutes the bulk of Internet traffic today. The performance criterion relevant to such transfers is the overall

document transfer time. This time is clearly highly dependent on the number of ongoing transfers on the links shared by the considered connection. This number varies as a random process as new connections are established and existing ones terminate in a way which depends on how bandwidth is allocated, as well as on the underlying traffic parameters.

In the case of a single bottleneck resource shared perfectly fairly, simple traffic assumptions of Poisson arrivals and identically and independently distributed document size lead to a processor sharing queueing model [5]. This fluid flow model provides useful results on expected response times as a function of the load of an access link or a Web server, for instance. It also shows how a form of congestion collapse can occur when demand (arrival rate \times mean document size) exceeds capacity. The processor sharing queue is then no longer ergodic leading to unbounded response times. To understand the impact of multiple bottlenecks and to investigate the effect of different sharing strategies, one would like to dispose of similar analytical results for multiple resource systems.

To the best of our knowledge, the only analytical results available so far are in Massoulié and Roberts [5], where the so-called linear network topology is investigated. Simulation results for the linear network can be found both in [5] and in de Veciana et al. [6]. The main motivation for the present paper is to study the performance of best-effort networks with alternative topologies, the ultimate objective being the derivation of heuristics enabling the performance evaluation of bandwidth sharing in a general network.

In the present paper we report the results of our preliminary investigations. These include an analysis of the stability conditions under which the expected response time remains finite in a general network. We also apply mean field techniques to evaluate the performance of large symmetrical networks. Numerical results derived from the model illustrate how response times depend on the number of bottleneck links and their utilization. These results are of some practical interest and aide our understanding of the behavior of best effort networks. A further significant contribution is the insight provided into the inherent difficulty of deriving performance estimates when more than one bottleneck limits throughput.

Section II introduces a general class of Markov models for best-effort networks which is intended to capture the impact of network topology, traffic parameters and bandwidth sharing (fairness) criteria on document transfer times. A brief account of the results obtained in [5] is given, and the so-called “min” bandwidth allocation is introduced as a conservative approximation to max-min fairness. Section III then establishes the necessary and sufficient ergodicity criteria for best-effort networks under the “min” policy. Section IV introduces the so-called “star topology”. Its relevance as a model of real networks is discussed

This work has been partly supported by a grant from France Télécom R&D.

G. Fayolle, A. de La Fortelle and J.-M. Lasgouttes are with INRIA, Domaine de Voluceau BP 105, Rocquencourt 78153 Le Chesnay CEDEX, France.

L. Massoulié is with Microsoft Research, Saint George House, 1 Guildhall Street, CB2 3NH Cambridge, United Kingdom.

J. Roberts is with France Télécom R&D, 38–40, rue du Général Leclerc, 92794 Issy les Moulineaux CEDEX 9, France.

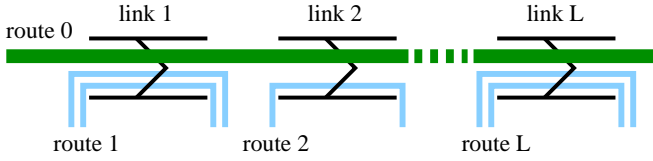


Fig. 1
THE LINEAR NETWORK

and a mean field heuristic is proposed. This heuristic is expected to be accurate in the asymptotic regime where the number of star branches is large. The derived fixed point equations are investigated numerically in Section V. Simulation is used to verify the accuracy of the heuristics. Extensions to the star-shaped network are also considered in Sections IV and V, which notably allow an evaluation of the impact of the number of bottlenecks on the mean transfer time.

II. BEST-EFFORT NETWORKS

Consider the following network model: a set \mathcal{L} of links is given, where each link $\ell \in \mathcal{L}$ has an associated capacity or bandwidth $C_\ell > 0$. A set \mathcal{R} of routes is given, each route being identified with a subset of links. Fig. 1 illustrates the so-called linear network: it consists of L links with equal capacity, route 0 which crosses each link, and routes $r = 1, \dots, L$ which cross a single link.

To each route r are associated two parameters: λ_r is the arrival rate of new transfer requests along route r , and σ_r is the average document size. We make the following standard simplifying assumptions: requests for document transfers along route r arrive at the instants of a Poisson process with intensity λ_r , while the corresponding document sizes are mutually independent, independent of the arrival times, and drawn from an exponential distribution of mean σ_r .

These traffic assumptions make the process specifying the number of transfers in progress on different routes Markovian (see below) and thus greatly simplify analysis. The Poisson arrivals assumption is not unreasonable in a large network. In view of the insensitivity of performance results for an isolated link to the exact document size distribution, we do not expect divergence of the real distribution from the exponential size assumption to invalidate the derived conclusions. However, the main reason for assuming an exponential distribution is clearly one of analytical tractability.

The network state is summarized by the variables $X \stackrel{\text{def}}{=} \{x_r, r \in \mathcal{R}\}$, where x_r denotes the number of transfers in progress along route r . It remains to specify at what speed documents are transmitted in any given state X in order to turn X into a Markov process with well defined dynamics. Indeed, given the rate $\zeta_r(X)$ at which documents along route r are transferred when the network state is X , X is a Markov process with non-zero transition rates given by

$$\begin{cases} x_r \rightarrow x_r + 1 : & \text{rate } \lambda_r, \\ x_r \rightarrow x_r - 1 : & \text{rate } x_r \zeta_r(X) / \sigma_r. \end{cases}$$

A natural assumption would be to consider that each document

receives its fair share of bandwidth. For instance, if as in [1] fairness is understood as max-min fairness, each transfer along route r receives a bandwidth share ζ_r^{mm} , where

$$\sum_{r \ni \ell} x_r \zeta_r^{\text{mm}} \leq C_\ell, \forall \ell \in \mathcal{L}, \quad (1)$$

and for every route r , there is at least one link $\ell \in r$ such that

$$\sum_{r' \ni \ell} x_{r'} \zeta_{r'}^{\text{mm}} = C_\ell, \text{ and } \zeta_r^{\text{mm}} = \max_{r' \ni \ell} \zeta_{r'}^{\text{mm}}. \quad (2)$$

These two conditions uniquely determine the bandwidth shares ζ_r^{mm} . Having specified the Markov process X , one can then attempt to study its steady state properties, identifying the conditions on the load parameters

$$\rho_\ell \stackrel{\text{def}}{=} \frac{1}{C_\ell} \sum_{r \ni \ell} \lambda_r \sigma_r$$

under which it is ergodic and, when it is, determining the stationary distribution. Mean transfer times T_r along each route r can then be computed using Little's law: $T_r = \mathbb{E}x_r / \lambda_r$.

It turns out that explicit formulas for steady state distributions are typically beyond reach. A notable exception is the linear network, with bandwidth shares being allocated to realize proportional rather than max-min fairness; see [5]. In order to obtain formulas in other cases, one therefore has to resort to asymptotics on various parameters. For instance, for the linear network with max-min fair rate sharing, the regime where the arrival rate λ_0 along route 0 goes to zero (essentially, a form of light traffic analysis) is considered in [5]; this leads to approximate formulas for T_0 . It can be shown, in particular, that T_0 increases as the logarithm of the number of links L when L increases. This is in contrast to the case of proportionally fair sharing where it increases linearly in L .

The main purpose of this paper is to investigate an alternative asymptotic regime where it is the network topology which evolves. The precise description of this limiting regime will be given in Section IV.

In the following sections, we consider bandwidth allocations according to the following “min” policy: given the network state X , each transfer along route r receives a bandwidth share ζ_r^{min} given by

$$\zeta_r^{\text{min}} \stackrel{\text{def}}{=} \min_{\ell \in r} \frac{C_\ell}{X_\ell}, \quad (3)$$

where we have introduced the notation $X_\ell = \sum_{r' \ni \ell} x_{r'}$ to represent the total number of transfers making use of link ℓ .

It is easy to check that this policy satisfies the capacity constraints (1). Moreover, it is sub-optimal with respect to the max-min fairness policy, as shown in the next theorem.

Theorem 1: Under the same initial conditions, the vector $X^{\text{mm}}(t)$ for the system under the ζ^{mm} allocation policy is stochastically smaller than $X^{\text{min}}(t)$, corresponding to the ζ^{min} allocation.

Proof: Assume that, for some t , $X^{\text{mm}}(t) \leq X^{\text{min}}(t)$. Then,

with the notation of (2),

$$\begin{aligned}\zeta_r^{\text{mm}} &= \max_{r' \ni \ell} \zeta_{r'}^{\text{mm}} \\ &\geq \frac{1}{X_\ell^{\text{mm}}(t)} \sum_{r' \ni \ell} x_{r'}^{\text{mm}}(t) \zeta_{r'}^{\text{mm}} = \frac{C_\ell}{X_\ell^{\text{mm}}(t)} \\ &\geq \frac{C_\ell}{X_\ell^{\text{min}}(t)} \geq \zeta_r^{\text{min}}.\end{aligned}$$

Thus, using a coupling argument, one can define the processes X^{mm} and $X^{\text{min}}(t)$ in such a way that $X^{\text{mm}}(t) \leq X^{\text{min}}(t)$ for all $t > 0$. ■

The previous theorem motivates the study of the min policy, as it implies for instance that mean transfer times T_r under the min policy provide upper bounds on the corresponding transfer times under the max-min policy.

III. ERGODICITY CONDITIONS

In the following we demonstrate that min and max-min bandwidth sharing policies have a stationary regime under the usual conditions, i.e. when the load on each link ℓ is less than 1:

Theorem 2: Under the allocation policies ζ^{mm} and ζ^{min} , the network is

- (i) ergodic if $\max_{\ell \in \mathcal{L}} \rho_\ell < 1$;
- (ii) transient if $\max_{\ell \in \mathcal{L}} \rho_\ell > 1$.

This result has already been proven for the max-min policy in [6]; we note that, by Theorem 1, ergodicity under the min policy implies ergodicity under the max-min policy, thus the treatment of the min policy given below provides an alternative proof to that of [6]. However we feel that, since the proof below is simpler and uses only elementary Lyapunov functions results, it should be easier to adapt to a more complicated situation. Transience under condition (ii) is in fact valid for any allocation policy which meets the capacity constraints (1).

Proof: Consider the discrete time chain $(\hat{X}(n), n \in \mathbb{N})$ describing the sequence of states visited by the continuous time jump process X . Transitions from a given state $\hat{X} = (\hat{x}_r, r \in \mathcal{R})$ satisfy

$$\begin{aligned}\mathbb{P}[\Delta \hat{x}_r(n) = 1 \mid \hat{X}(n) = \hat{X}] &= \frac{\lambda_r}{D}, \\ \mathbb{P}[\Delta \hat{x}_r(n) = -1 \mid \hat{X}(n) = \hat{X}] &= \frac{1}{D} \frac{\hat{x}_r}{\sigma_r} \min_{\ell \in r} \frac{C_\ell}{\hat{X}_\ell},\end{aligned}$$

where $\Delta \hat{x}_r(n) \stackrel{\text{def}}{=} \hat{x}_r(n+1) - \hat{x}_r(n)$ and

$$\begin{aligned}D &\stackrel{\text{def}}{=} \sum_{r \in \mathcal{R}} \left(\lambda_r + \frac{\hat{x}_r}{\sigma_r} \min_{\ell \in r} \frac{C_\ell}{\hat{X}_\ell} \right) \\ &\leq |\mathcal{R}| \cdot \left(\max_{r \in \mathcal{R}} \lambda_r + \max_{r \in \mathcal{R}} \frac{1}{\sigma_r} \cdot \max_{\ell \in \mathcal{L}} C_\ell \right) \\ &\stackrel{\text{def}}{=} D'.\end{aligned}$$

Ergodicity of the continuous time process X will follow from that of \hat{X} and from the fact that the mean sojourn times in each state X are bounded from above uniformly in X (or equivalently, that the jump rates out of each state X are bounded away from zero), a property which is easily verified.

Sufficient condition. Assume $\rho_M \stackrel{\text{def}}{=} \max_{\ell \in \mathcal{L}} \rho_\ell < 1$, and define the Lyapunov function

$$f(\hat{X}) \stackrel{\text{def}}{=} \sum_{r \in \mathcal{R}} \sum_{1 \leq k \leq \hat{x}_r} \gamma_r^k,$$

where $\gamma_r > 1$ will be chosen later. The structure of the function, which may seem unnatural, has been chosen for the sake of computation; it is in fact of the general form $\sum_{r \in \mathcal{R}} \beta_r \gamma_r^{\hat{x}_r} + K$, for appropriate constants β_r and K .

In order to express the transition rates in terms of \hat{x}_r and ρ_ℓ , remark that

$$\hat{X}_\ell = \sum_{r \ni \ell} \hat{x}_r = \sum_{r \ni \ell} \lambda_r \sigma_r \frac{\hat{x}_r}{\lambda_r \sigma_r} \leq \rho_\ell C_\ell \max_{r \ni \ell} \frac{\hat{x}_r}{\lambda_r \sigma_r}.$$

Then, using the notation

$$\hat{x}_r^* \stackrel{\text{def}}{=} \frac{\hat{x}_r}{\lambda_r \sigma_r}, \quad \hat{x}_M^* \stackrel{\text{def}}{=} \max_{r \in \mathcal{R}} \hat{x}_r^*,$$

we have, $\forall r \in \mathcal{R}$,

$$\mathbb{P}[\Delta \hat{x}_r(n) = -1 \mid \hat{X}(n) = \hat{X}] \geq \frac{\lambda_r}{D \rho_M} \frac{\hat{x}_r^*}{\hat{x}_M^*}.$$

Thus,

$$\begin{aligned}\mathbb{E}[f(\hat{X}(n+1)) - f(\hat{X}(n)) \mid \hat{X}(n) = \hat{X}] \\ &= \sum_{r \in \mathcal{R}} \left(\gamma_r^{\hat{x}_r+1} \mathbb{P}[\Delta \hat{x}_r(n) = 1 \mid \hat{X}(n) = \hat{X}] \right. \\ &\quad \left. - \gamma_r^{\hat{x}_r} \mathbb{P}[\Delta \hat{x}_r(n) = -1 \mid \hat{X}(n) = \hat{X}] \right) \\ &\leq \sum_{r \in \mathcal{R}} \frac{\lambda_r}{\rho_M D} \gamma_r^{\hat{x}_r} \left[\rho_M \gamma_r - \frac{\hat{x}_r^*}{\hat{x}_M^*} \right]\end{aligned}$$

Let $\gamma_r \stackrel{\text{def}}{=} \gamma^{\frac{1}{\lambda_r \sigma_r}}$, where γ is such that

$$\rho_M \gamma_r = \rho_M \gamma^{\frac{1}{\lambda_r \sigma_r}} < \theta < 1, \quad r \in \mathcal{R},$$

for some real number θ satisfying $\rho_M < \theta < 1$. The following inequality sums up what we have so far:

$$\begin{aligned}\mathbb{E}[f(\hat{X}(n+1)) - f(\hat{X}(n)) \mid \hat{X}(n) = \hat{X}] \\ \leq \sum_{r \in \mathcal{R}} \frac{\lambda_r \gamma_r^{\hat{x}_r}}{\rho_M D} \left[\theta - \frac{\hat{x}_r^*}{\hat{x}_M^*} \right].\end{aligned}$$

Let α be a real number such that $\theta < \alpha < 1$. The following quantities will be evaluated separately

$$\begin{cases} \Sigma_1 \stackrel{\text{def}}{=} \sum_{r: \hat{x}_r^* > \alpha \hat{x}_M^*} \frac{\lambda_r \gamma_r^{\hat{x}_r}}{\rho_M D} \left[\theta - \frac{\hat{x}_r^*}{\hat{x}_M^*} \right], \\ \Sigma_2 \stackrel{\text{def}}{=} \sum_{r: \hat{x}_r^* \leq \alpha \hat{x}_M^*} \frac{\lambda_r \gamma_r^{\hat{x}_r}}{\rho_M D} \left[\theta - \frac{\hat{x}_r^*}{\hat{x}_M^*} \right]. \end{cases}$$

Since Σ_1 is a sum of negative terms, the following bound holds, for any r_0 is such that $\hat{x}_{r_0}^* = \hat{x}_M^*$,

$$\Sigma_1 \leq \frac{\lambda_{r_0} \gamma_{r_0}^{\hat{x}_M^*}}{\rho_M D} (\theta - \alpha) \leq \frac{\gamma_M^{\hat{x}_M^*}}{\rho_M D} (\theta - \alpha) \min_{r \in \mathcal{R}} \lambda_r < 0.$$

Bounding Σ_2 is straightforward:

$$\Sigma_2 \leq \sum_{r: \hat{x}_r^* \leq \alpha \hat{x}_M^*} \frac{\lambda_r \gamma^{\alpha \hat{x}_M^*}}{\rho_M D} \theta \leq \frac{\gamma_M^{\hat{x}_M^*}}{\rho_M D} \gamma^{(\alpha-1)\hat{x}_M^*} |\mathcal{R}| \theta \max_{r \in \mathcal{R}} \lambda_r.$$

Now, if $C > 0$ and $\varepsilon > 0$ are chosen to satisfy the inequality

$$(\theta - \alpha) \min_{r \in \mathcal{R}} \lambda_r + \gamma^{(\alpha-1)C} |\mathcal{R}| \theta \max_{r \in \mathcal{R}} \lambda_r \leq -\varepsilon,$$

we have, $\forall \hat{X} \in \{\hat{x}_M^* > C\}$,

$$\begin{aligned} \mathbb{E}[f(\hat{X}(n+1)) - f(\hat{X}(n)) \mid \hat{X}(n) = \hat{X}] \\ = \Sigma_1 + \Sigma_2 \leq -\varepsilon \frac{\gamma_M^{\hat{x}_M^*}}{\rho_M D} \leq -\varepsilon \frac{\gamma^C}{\rho_M D'} < 0. \end{aligned}$$

Since $\{\hat{x}_M^* \leq C\}$ is a compact set, Foster's theorem applies (see e.g. [7]) and the Markov chain is ergodic.

Necessary condition. Assume now that there exists ℓ_0 such that $\rho_{\ell_0} > 1$. Defining

$$g(\hat{X}) \stackrel{\text{def}}{=} \sum_{r \geq \ell_0} \sigma_r \hat{x}_r,$$

we immediately have,

$$\begin{aligned} \mathbb{E}[g(\hat{X}(n+1)) - g(\hat{X}(n)) \mid \hat{X}(n) = \hat{X}] \\ = \sum_{r \geq \ell_0} \sigma_r \left(\mathbb{P}[\Delta \hat{x}_r(n) = 1 \mid \hat{X}(n) = \hat{X}] \right. \\ \left. - \mathbb{P}[\Delta \hat{x}_r(n) = -1 \mid \hat{X}(n) = \hat{X}] \right) \\ \geq \frac{1}{D'} [C_{\ell_0} \rho_{\ell_0} - C_{\ell_0}] > 0. \end{aligned}$$

Since the jumps are bounded, the chain is transient. \blacksquare

IV. MEAN FIELD ANALYSIS OF LARGE NETWORKS

It does not appear possible to obtain closed form expressions for the stationary distribution of the best-effort network state under the min policy. We therefore turn to the study of these stationary distributions under a limiting regime on network size and topology. A similar approach has previously been successfully applied to loss networks (see [8], [9], and references therein), and to queueing networks in [10], [11]. It is inspired by the so-called mean field models of statistical physics.

Mean field analysis in the present context is best illustrated by the star-shaped network of Fig. 2. This network has $N/2$ branches, each consisting of one inbound and one outbound link (thus implicitly N is an even number), and all links have unit capacity. Each route r connects the endpoints of two branches via the center node. It has an associated arrival rate $2\lambda/N$ and mean message length σ . The factor $2/N$ is introduced to make the total load on each link, $\rho = \lambda\sigma$, independent of the number of links N . As discussed below, when N goes to infinity the number of ongoing transfers on any link becomes independent of the number of ongoing transfers on any finite collection of the other links (this was termed the ‘‘chaos propagation’’ property in [9]). This allows us to derive fixed point equations for the probability distribution of the number of transfers in progress on any link.

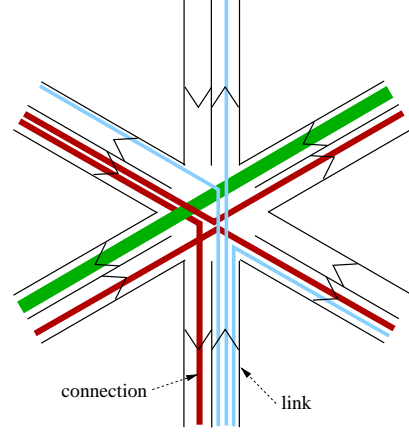


Fig. 2

A STAR-SHAPED NETWORK

A. Symmetrical star-shaped networks

Although amenability to a mean field analysis is a significant motivation for considering the star-shaped topology, it should be noted that it is also relevant to the study of real networks. Any overprovisioned links in a real network are largely transparent to the throughput of elastic best effort flows. Only bottleneck resources, typically located in the access network and within Web servers, need to be included in the network model. The star-shaped network may thus be considered to represent any network with a well provisioned backbone where throughput is limited by bottlenecks at the source and destination edges. For example, inbound links might represent Web server CPU, while outbound links correspond to the last hop of an ISP's interconnection network. This discussion not only motivates the consideration of such a topology, but also suggests that letting N go to infinity might indeed be realistic if N represents the number of Web servers over the Internet. Of course, there would be no reason in practice to assume symmetry. This assumption is introduced solely for reasons of tractability.

Although our focus is on the star-shaped topology, the mean field approach can be applied to other symmetrical topologies. It thus allows one to consider routes with more than two hops. The corresponding extended model is described in detail in Section IV-B below, where the corresponding fixed point equations are derived. Section IV-C then presents analytical results for the star-shaped network. Results of numerical investigation of the fixed point equations are reported in Section V.

B. Fixed point equations for large symmetrical networks

We use the following notation in the sequel.

- N : total number of links;
- L : length of a route through the network;
- $R^{(N)}$: number of routes going through a given link;
- $X_\ell^{(N)}$: number of active connections on link $\ell \in \mathcal{L}$, in stationary state;
- $x_r^{(N)}$: number of active connections on route r , going through links $r(1), \dots, r(L)$;
- λ : arrival rate on a link;

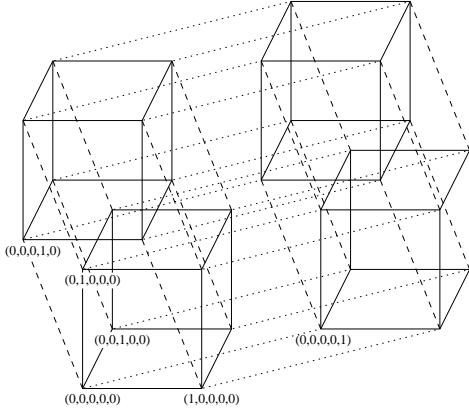


Fig. 3

A HYPERCUBE-SHAPED NETWORK WITH DIMENSION 5.

- σ : mean message length;
- $\rho \stackrel{\text{def}}{=} \lambda\sigma$: load of a link.

We have implicitly assumed here that the number of routes going through a link, $R^{(N)}$, is the same for all links. We shall in fact assume further that the network topology is the same, as seen from any route. We do not attempt to give a formal definition of this symmetry assumption here. The reader is referred to [9] for a thorough discussion on the minimal symmetry assumptions required. Symmetry implies notably that each route has the same number of hops and the same traffic parameters. The star-shaped network discussed above constitutes an example of such a symmetrical network when $L = 2$ (with $R^{(N)} = N/2$).

It is more difficult to come up with meaningful examples of symmetrical networks supporting routes with $L > 2$: in particular, the network should not be fully connected, since routes longer than one hop would then be pointless. One reasonable model is a hypercube (Fig. 3) of large dimension, in which each edge contains two one-way links.

The hypercube is a classical structure with many symmetries. It is characterized by its dimension d . Its vertices are represented by d -tuples of 0s and 1s (e.g., $(0, 1, 1, 0, 1)$) and its edges connect two vertices differing in only one coordinate.

The total number of links in such a network is $N = d2^d$. The number of routes going through any link is

$$R^{(N)} = L \frac{(d-1)!}{(d-L)!},$$

where the only routes considered are the shortest paths between two vertices which differ in exactly L coordinates. Note that the results below do not depend on the precise topology of the network.

We now derive the fixed point equations. It should be stressed that this derivation is heuristic. We clearly mention which steps need further justification in the course of the derivation. We do believe that the equations are very good approximations, however, especially in view of the numerical and simulation results presented in the following section.

Assume now that $\rho < 1$ and that the system is in stationary state $X^{(N)} = (x_r^{(N)}, r \in \mathcal{R})$. For any $k \geq 0$, the proportion of links in state k is

$$\alpha_k^{(N)} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\ell \in \mathcal{L}} \mathbb{1}_{\{X_\ell^{(N)} = k\}}.$$

By symmetry, it holds that

$$\mathbb{P}(X_\ell^{(N)} = k) = \mathbb{E} \alpha_k^{(N)}, \forall \ell \in \mathcal{L}.$$

The chaos propagation assumption¹ implies that $\alpha_k^{(N)}$ obeys a law of large numbers:

$$\alpha_k \stackrel{\text{def}}{=} \lim_{N \rightarrow \infty} \alpha_k^{(N)} = \lim_{N \rightarrow \infty} \mathbb{P}(X_\ell^{(N)} = k) \stackrel{\text{def}}{=} \mathbb{P}(X = k).$$

It appears that the dynamics of the system are driven by $\alpha_k^{(N)}$, traditionally referred to as the *mean field*. The following notation will also be useful:

$$\bar{\alpha}^{(N)} \stackrel{\text{def}}{=} \sum_{k \geq 0} k \alpha_k^{(N)}, \quad \bar{\alpha} \stackrel{\text{def}}{=} \mathbb{E} X_\ell = \sum_{k \geq 0} k \alpha_k.$$

In order to derive the equation satisfied by the limit stationary distribution α_k , we must first describe the possible transitions for $\alpha_k^{(N)}$. The two cases of interest are

- arrival on a link with k connections:

$$\alpha_k^{(N)} \rightarrow \alpha_k^{(N)} - \frac{1}{N}, \quad \alpha_{k+1}^{(N)} \rightarrow \alpha_{k+1}^{(N)} + \frac{1}{N}$$

- departure from a link with $k > 0$ connections:

$$\alpha_k^{(N)} \rightarrow \alpha_k^{(N)} - \frac{1}{N}, \quad \alpha_{k-1}^{(N)} \rightarrow \alpha_{k-1}^{(N)} + \frac{1}{N}$$

The transition corresponding to a new connection arrival has rate λ . The main problem is to compute the departure rate from a link ℓ , given that it has $X_\ell^{(N)} = k$ ongoing transfers. This can be written as

$$\frac{1}{\sigma} \sum_{r \ni \ell} \mathbb{E} \left[x_r^{(N)} \min_{\ell' \in r} \frac{1}{X_{r(\ell')}^{(N)}} \mid X_\ell^{(N)} = k \right]. \quad (4)$$

Since the total number of routes is much larger than the size N of the network, we assume that the probability of having more than one connection on a route r is negligible², and that the links on route r are independent, conditioned on $\{x_r^{(N)} = 1\}$ ³. The first property allows to rewrite (4) as

$$\frac{1}{\sigma} \sum_{r \ni \ell} \mathbb{E} [x_r^{(N)} \mid X_\ell^{(N)} = k] \mathbb{E} \left[\min_{\ell' \in r} \frac{1}{X_{r(\ell')}^{(N)}} \mid x_r^{(N)} = 1 \right]. \quad (5)$$

Let j be a given link and let r be one route using link j , i.e., $r \ni j$. The distribution of $X_j^{(N)}$, conditioned on there being one connection on r , is then

$$\mathbb{P}[X_j^{(N)} = k' \mid x_r^{(N)} = 1] = \frac{\mathbb{E} [\mathbb{1}_{\{X_j^{(N)} = k'\}} \mathbb{1}_{\{x_r^{(N)} = 1\}}]}{\mathbb{E} \mathbb{1}_{\{x_r^{(N)} = 1\}}}.$$

¹We have not proven that this assumption holds. It seems, however, that the techniques developed in [9] could be applied to prove that this is the case, provided the parameter $R^{(N)}$ goes to infinity with N .

²This fact is easy to prove in finite time, but requires more work for the stationary regime.

³This is the point where the heuristic is not completely exact; it is however likely to be true when ρ tends either to 0 or 1.

By symmetry, it is possible to sum both sides of the above fraction over all the routes going through link j ,

$$\begin{aligned}\mathbb{P}[X_j^{(N)} = k' \mid x_r^{(N)} = 1] &= \frac{\mathbb{E}[X_j^{(N)} \mathbb{1}_{\{X_j^{(N)} = k'\}}]}{\mathbb{E}X_j^{(N)}} \\ &= \frac{k' \mathbb{E}\alpha_{k'}^{(N)}}{\mathbb{E}\bar{\alpha}^{(N)}}.\end{aligned}$$

Departure rate (5) then becomes, in view of the assumed independence property between the $X_{r(i)}$ given $x_r = 1$,

$$\frac{k}{\sigma} \sum_{k_2, \dots, k_L=1}^{\infty} \left(\frac{1}{k} \wedge \frac{1}{k_2} \wedge \dots \wedge \frac{1}{k_L} \right) \prod_{i=2}^L \frac{k_i \mathbb{E}\alpha_{k_i}^{(N)}}{\mathbb{E}\bar{\alpha}^{(N)}}.$$

Taking the limit $N \rightarrow \infty$, the invariant measure equations follow. We have:

$$-\lambda\alpha_0 + \frac{1}{\sigma}\alpha_1 \frac{u_1}{\bar{\alpha}^{L-1}} = 0 \quad (6)$$

and

$$\lambda(\alpha_{k-1} - \alpha_k) + \frac{1}{\sigma} \left(\alpha_{k+1} \frac{u_{k+1}}{\bar{\alpha}^{L-1}} - \alpha_k \frac{u_k}{\bar{\alpha}^{L-1}} \right) = 0 \quad (7)$$

for $k \geq 1$, where

$$u_k \stackrel{\text{def}}{=} k \sum_{k_2, \dots, k_L=1}^{\infty} \left(\frac{1}{k} \wedge \frac{1}{k_2} \wedge \dots \wedge \frac{1}{k_L} \right) \prod_{i=2}^L k_i \alpha_{k_i}. \quad (8)$$

Equations (6) and (7) can be rewritten in a more concise form as

$$\alpha_{k+1}u_{k+1} = \rho\bar{\alpha}^{L-1}\alpha_k, \quad \forall k \geq 0. \quad (9)$$

The two sets of equations (8) and (9) together constitute the fixed point equations we require. As noted in the introduction of this section, these equations do not depend on the topology of the network. The expression for u_k can be simplified.

Let Y_2, \dots, Y_L be random variables with distribution

$$\mathbb{P}(Y_i = y) = \frac{y\alpha_y}{\bar{\alpha}}, \quad 2 \leq i \leq L,$$

and let $Y \stackrel{\text{def}}{=} \max(Y_2, \dots, Y_L)$. Then (8) reads

$$u_k = k\bar{\alpha}^{L-1} \mathbb{E}\left[\frac{1}{k} \wedge \frac{1}{Y}\right]. \quad (10)$$

Straightforward calculations yield

$$\begin{aligned}\mathbb{E}\left[\frac{1}{k} \wedge \frac{1}{Y}\right] &= \sum_{y=1}^{\infty} \left[\frac{1}{k} \wedge \frac{1}{y}\right] \mathbb{P}(Y = y) \\ &= \sum_{y=k}^{\infty} \frac{1}{y(y+1)} \mathbb{P}(Y \leq y).\end{aligned}$$

The simplified form for u_k is thus, from the basic properties of the minimum of independent random variables:

$$u_k = k \sum_{y=k}^{\infty} \frac{1}{y(y+1)} \left[\sum_{m=0}^y m\alpha_m \right]^{L-1}. \quad (11)$$

Note that, in the case $L = 2$ (i.e., for the star-shaped network), the original equation (8) is perhaps simpler than the equivalent expression (11). It yields the following form for the fixed point equations:

$$\alpha_{k+1}u_{k+1} = \rho\bar{\alpha}\alpha_k, \quad (12)$$

$$u_k = \sum_{y>0} (k \wedge y) \alpha_y. \quad (13)$$

Remark 1: When considering the star-shaped network as a model for Web transfers over the Internet, as suggested in Section IV-A, inbound links could be seen as the CPU of Web servers and outbound links as the last hop between the ISP's backbone and the end customers. It thus makes sense to relax the symmetry assumption we had made between inbound and outbound links, as the two types of bottlenecks are of a different nature. We might thus consider a star-shaped network with N^{in} inbound links, N^{out} outbound links, inbound (resp. outbound) links having capacity C^{in} (resp. C^{out}), see Fig. 4. Assume the mean message length σ is the same for each two-hop route, and the link capacities C^{out} are fixed. The arrival rate on each route has the form $\lambda_r \stackrel{\text{def}}{=} \lambda/N^{\text{in}}$, and the load $\rho^{\text{out}} \stackrel{\text{def}}{=} \lambda\sigma/C^{\text{out}}$ is less than 1. The capacity C^{in} of a “backbone” link is chosen to ensure $\rho^{\text{in}} \stackrel{\text{def}}{=} \lambda\sigma N^{\text{out}}/C^{\text{in}} N^{\text{in}} < 1$ is fixed when the size of the system grows. Then, as N^{in} and N^{out} increase, with $N^{\text{in}}/N^{\text{out}}$ small, the inbound links have many active connections and a large capacity, while the outbound links remain in “normal” utilization. The same approach as above can then be applied, to yield the set of fixed point equations

$$\begin{aligned}\alpha_{k+1}^{\text{in}} u_{k+1}^{\text{out}} &= \rho^{\text{in}} \bar{\alpha}^{\text{out}} \alpha_k^{\text{in}}, \\ \alpha_{k+1}^{\text{out}} u_{k+1}^{\text{in}} &= \rho^{\text{in}} \bar{\alpha}^{\text{out}} \alpha_k^{\text{out}},\end{aligned}$$

where

$$\begin{aligned}u_k^{\text{in}} &= \sum_{y>0} \alpha_y^{\text{in}} \min\left[k, \frac{C^{\text{out}}}{C^{\text{in}}} y\right], \\ u_k^{\text{out}} &= \sum_{y>0} \alpha_y^{\text{out}} \min\left[\frac{C^{\text{out}}}{C^{\text{in}}} k, y\right],\end{aligned}$$

and α_k^{in} (resp. α_k^{out}) represents the proportion of inbound (resp. outbound) links with k ongoing transfers.

C. Analytical results for $L = 2$

While equation (9) looks superficially like a “birth and death process” equation, it is in fact non-linear due to the fact that u_k and $\bar{\alpha}$ both depend on the α_y .

From (10), one clearly sees that u_k is increasing in k , and tends to $\bar{\alpha}^{L-1}$ when $k \rightarrow \infty$. Therefore, α_k is increasing as long as $u_k < \rho\bar{\alpha}^{L-1}$, and decreasing after that. This means that the α_k form a modal distribution, which maximal value is attained at

$$k_0 = \max\{k > 0, u_k < \rho\bar{\alpha}^{L-1}\}.$$

We now present analytical results on the solution of the fixed point equations for $L = 2$. The proof of these results can be found in [12]. It relies heavily on functional analysis.

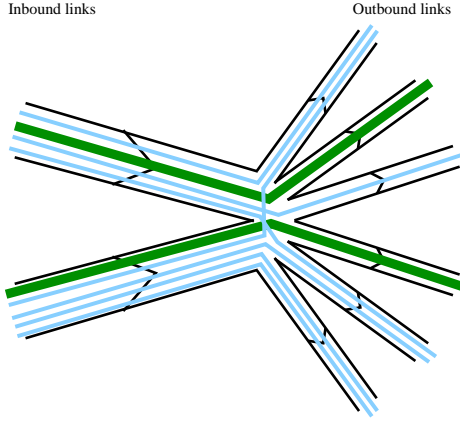


Fig. 4

THE ASYMMETRICAL STAR-SHAPED NETWORK

Equations (12)–(13) have an unique solution for $\rho < 1$ and, when $\rho \rightarrow 1$, the following asymptotic expansions hold:

$$\begin{aligned} \mathbb{E}X &\approx \frac{1}{(1-\rho)^2 A}, \\ \lim_{k \rightarrow \infty} \rho^{-k} \mathbb{P}(X = k) &\approx (1-\rho) B \exp\left[\frac{1}{(1-\rho)A}\right], \end{aligned}$$

where A and B are non-negative constants.

Thus, under the min policy, any link in the star-shaped network has a mean queue length which is one order of magnitude larger than for a single server queue with the same load ($\rho/(1-\rho)$). Its tail distribution is still geometrical with factor ρ .

It is possible to give an expression for the constant A : if c and v are solutions of the following system of differential equations,

$$\begin{cases} zc'(z) + c(z)v(z) = 0, \\ zv''(z) + v'(z) = c(z), \\ v(0) = 0, \quad v'(0) = 1, \quad c(0) = 1, \end{cases}$$

then A can be written as follows:

$$A = \int_0^\infty c(z) dz = \lim_{z \rightarrow \infty} zv'(z) \approx 1.30.$$

Since this system is numerically highly unstable, it has proven difficult (with the “Livermore stiff ODE” solver from MAPLE) to derive a better estimate for A .

It is interesting to note that the function $w(y) \stackrel{\text{def}}{=} v(e^y) + 1$ satisfies the so-called Blasius [13] equation

$$w'''(y) + w(y)w''(y) = 0,$$

which describes a laminar boundary layer along a flat plate (see, e.g., [14]).

V. NUMERICAL ANALYSIS AND SIMULATIONS

While the analytical results of Section IV-C give some good estimates, they are only valid in the heavy traffic regime $\rho \rightarrow 1$.

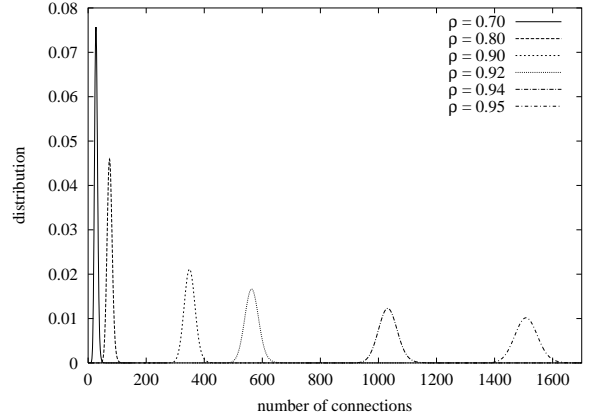


Fig. 5

DISTRIBUTION OF THE MEAN-FIELD PROBABILITIES $\mathbb{P}(X = k)$ FOR $L = 20$ AND DIFFERENT VALUES OF THE LOAD ρ .

In addition, similar results for $L > 2$ are not available. We thus resort to numerical resolution of the equations to gain a better understanding of the performance of transfers across large symmetrical networks. The very form of the equations suggests the use of a fixed-point method for this numerical resolution: starting from *a priori* values $(\alpha_k^{(0)}, k \geq 0)$, the algorithm computes the corresponding $u_k^{(0)}$ from (8), and then new values $\alpha_k^{(1)}$ from (9). Provided special care is taken to avoid instabilities, the iteration of this process converges rapidly (less than 100 steps). Sample results are shown in Fig. 5, corresponding to a large symmetrical network with routes of length $L = 20$.

As clearly seen in the figure, the distributions are very different from what would be obtained for routes of length $L = 1$. In this case, the system consists of a collection of independent M/M/1 queues and the associated distribution $\{\alpha_k\}$ is geometric. For $L > 1$, the $\{\alpha_k\}$ distributions are markedly modal and the positions of the peak values are roughly proportional to $(1-\rho)^{-2}$, a fact has only been proven in Section IV-C for the case $L = 2$. Moreover, since the shape of the distribution is rather narrow, this position roughly coincides with the mean number of active connection (as can be seen from the raw data).

The impact of route length is illustrated in Fig. 6. It seems that the mean number of active connections (which is again approximately the peak value of the distribution) is roughly proportional to $\log L$. Note that a logarithmic growth rate is very slow suggesting that, beyond 2 or 3, the number of bottlenecks does not have a significant impact on mean transfer times. They depend much more on the load ρ .

The results presented so far only concern the solution of the fixed point equations. As mentioned earlier, there are gaps in the derivation of these equations. To assess their quality and to investigate the accuracy of the asymptotic approximation for finite size networks, we ran a number of simulations of the star-shaped network. Fig. 7 displays the corresponding results when the load on each link is set to $\rho = 0.9$ for a varying number of links. The agreement between the simulation results and the fixed point equation results is excellent for $N = 100$ links and

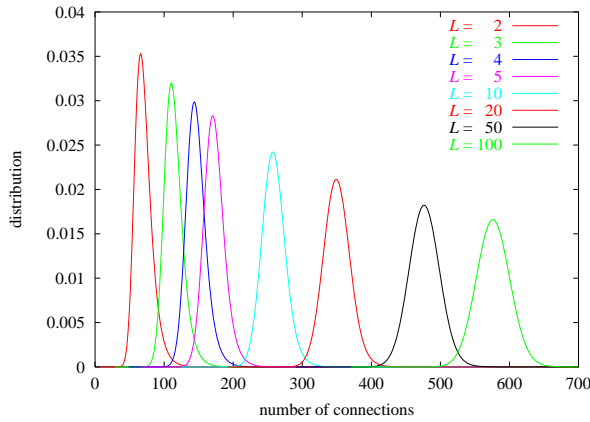


Fig. 6

DISTRIBUTION OF THE MEAN-FIELD PROBABILITIES $\mathbb{P}(X=k)$ FOR DIFFERENT VALUES OF THE ROUTE LENGTH L AND LOAD $\rho = 0.9$.

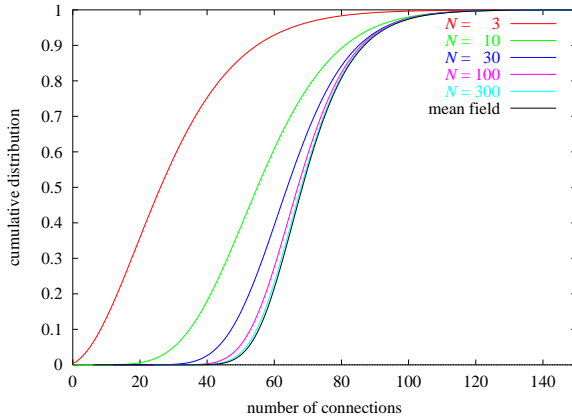


Fig. 7

CUMULATIVE DISTRIBUTION $\mathbb{P}(X \leq k)$ FOR DIFFERENT VALUES OF N (SIMULATION) AND INFINITE SIZE (FIXED POINT) FOR A STAR-SHAPED NETWORK WITH LOAD $\rho = 0.9$.

improves as N increases.

VI. CONCLUSIONS

In this paper, we have considered a class of Markov processes called best-effort networks which constitutes a natural probabilistic model for evaluating the performance of document transfers over data networks such as the Internet. Unlike almost all previous work, this model accounts for the random nature of traffic: document transfers begin at the epochs of a certain arrival process and the size of each document is drawn from a given probability distribution. In the interests of tractability we assumed Poisson arrivals and exponentially distributed sizes. We introduced the “min” bandwidth sharing policy as a conservative approximation to the more classical max-min policy. Necessary and sufficient ergodicity conditions for best effort networks under the min and max-min policies have been established.

In order to pursue the analysis of the stationary distributions of the number of transfers in progress, we have resorted to large network asymptotics applying the mean field approach of statistical physics. This enabled us to derive fixed point equations for the probability distribution of the number of ongoing transfers on a given network link. The validity of these equations has been established by comparing their solution with the results of simulations.

Analytical and numerical results show how the mean transfer time depends on the number of bottleneck links and their load. The steady state distribution in networks where routes have several bottlenecks ($L > 1$) has a marked modal behavior. This is significantly different to the geometric distribution which holds when routes have a single bottleneck ($L = 1$). Performance is also much more sensitive to link load ρ for multiple bottleneck routes: as $\rho \rightarrow 1$, mean transfer time increases like $1/(1-\rho)^2$ in the case $L = 2$, whereas the dependence is in $1/(1-\rho)$ when $L = 1$. Finally, the impact of the number of hops per route L appears small (given that $L > 1$) compared to that of parameter ρ . This suggests that the star-shaped network is perhaps a sufficiently complex model, and that the study of symmetrical networks with $L > 2$ is less relevant.

The work presented here can be pursued in several directions. On the theoretical side, the analytical results presented in Section IV-C constitute a first step to understanding the solution of the fixed point equations which could be taken further. Another challenging theoretical question is to improve the fixed point equations in a rigorous way. On a more practical side, the fixed point equations might be simplified so as to find simple approximate formulas for mean transfer times as a function of key parameters (such as ρ^{in} , ρ^{out} in the case of the asymmetrical star-shaped network described in Remark 1). Such approximate formulas could then lead to engineering rules for capacity planning.

We view the present study as a preliminary investigation into the performance of best effort networks with multiple bottleneck links. A significant result of this investigation is the discovery that the extension of the processor sharing model valid for a single bottleneck proves to be very hard. There appears to be no simple parallel to the familiar fixed point techniques used in loss networks. The problem is, however, of considerable practical importance for providers seeking to engineer their network to ensure adequate throughput for document transfers. We hope therefore that this paper will incite further work and the development of alternative heuristic approaches.

REFERENCES

- [1] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall International, 2nd edition, 1992.
- [2] J. H. Mo and J. Walrand, “Fair end-to-end window-based congestion control,” in *SPIE '98 International Symposium on Voice, Video and Data Communications*, 1998.
- [3] L. Massoulié and J. Roberts, “Bandwidth sharing: objectives and algorithms,” *IEEE Infocom*, 1999.
- [4] P. Hurley, J.-Y. Leboudec, and P. Thiran, “A note on the fairness of additive increase and multiplicative decrease,” *ITC 16*, 1999.
- [5] J. Roberts and L. Massoulié, “Bandwidth sharing and admission control for elastic traffic,” *ITC specialists seminar*, 1998.
- [6] G. De Veciana, T.-J. Lee, and T. Konstantopoulos, “Stability and performance analysis of networks supporting services with rate control – could the internet be unstable?,” *IEEE Infocom*, 1999.

- [7] G. Fayolle, V. A. Malyshev, and M. V. Menshikov, *Topics in the Constructive Theory of Countable Markov Chains*, Cambridge University Press, 1995.
- [8] F. Kelly, "Loss networks," *Ann. Appl. Probab.*, vol. 1, pp. 319–378, 1991.
- [9] C. Graham and S. Méléard, "Chaos hypothesis for a system interacting through shared resources," *Probability Theory and Related Fields*, 1994.
- [10] N. D. Vvedenskaya, R. L. Dobrushin, and F. I. Karpelevich, "A queueing system with a choice of the shorter of two queues—an asymptotic approach," *Problems Inform. Transmission*, vol. 32, pp. 15–27, 1996.
- [11] F. Delcoigne and G. Fayolle, "Thermodynamical limit and propagation of chaos in polling systems," *Markov Processes and Related Fields*, vol. 5, no. 1, pp. 89–124, 1999.
- [12] G. Fayolle and J.-M. Lasgouttes, "A nonlinear integral operator encountered in the bandwidth sharing of a star-shaped network," in *Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*, Trends in Mathematics, pp. 231–242. Birkhauser, 2000.
- [13] H. Blasius, "Grenzschichten in flüssigkeiten mit kleiner reibung," *Z. Math. Phys.*, vol. 56, pp. 1–37, 1908, English translation in NACA TM 1256.
- [14] H. Schlichting, *Boundary layer theory*, McGraw-Hill Book Co., Inc., New York, 1960, Translated by J. Kestin. 4th ed. McGraw-Hill Series in Mechanical Engineering.